
Data, Archives and Software

Lorella Angelini
NASA Goddard

Topics

- **Data format**
- **Archives**
- **CALDB**
- **Software**

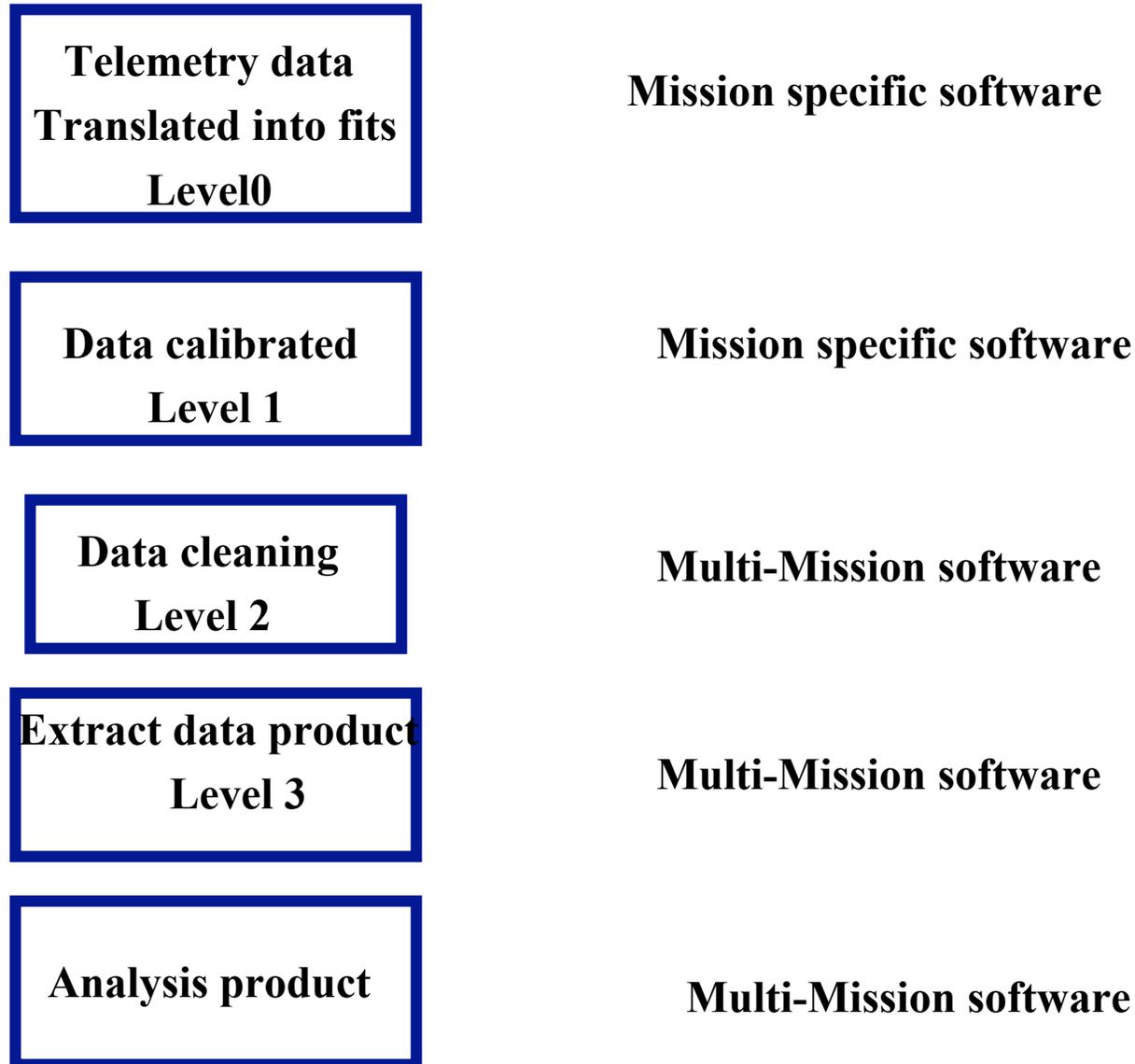
Look from the past

- Prior to 1990 it was custom to create ad-hoc data format, software with the calibration information typically embedded
- Archives did not exist or they were not accessible on-line
- Therefore it was quite difficult for astronomer not connected with the a specific experiment to work on specific data set
- The software in most cases was not portable
- The data format was not general making difficult to combine data that in nature were similar
- High energy astrophysics deals with data from satellite
 - These are ‘expensive’ data
 - Not easy to repeat or obtain

Current way of deal with data

- In 1990 several factors contributed to change the way now we deal with high energy astrophysics data
- FITS (Flexible Image Transport system) as data format :
 - Format used in optical astronomy but expanded to accommodate high energy astrophysics data
- Software :
 - Adopted a common user interface (Parameter file)
 - Separate the mission specific software from the generic/ general software
 - Mission specific software deal mostly with the specifics of the instrument and its calibration
 - Generic software deals with analysis that is common to all experiments, for example creating a lightcurve, image manipulation, spectral fitting
- Calibration data were decoupled in files and no longer included in the software (see CALDB)
- Data were put in archives and the archives became accessible on-line

From the telemetry to the archive



FITS file (1)

- Originally FITS was used to “transport data” in a platform independent way
- A FITS file comprises of
 - an ASCII Header where information is provided as Keyword= Value
 - and a binary array
- End on 1990 the FITS file format was upgraded to include the so called Bintable extension
- This is similar to a table with columns each containing different quantities
- X-ray and Gamma ray data were ideal to layout into a Bintable
 - Where a column represents an event attribute (position , energy etc..)
 - And a row gives all the attributes for that event or an histogram of for example energy integrated over a time period
- After the adoption of the Bintable several standards format were established
 - Selecting specific keywords to indicate special quantities, e.g. RA_OBJ
 - Creating rules of how layout a lightcurve a spectrum or a response matrix

Standards in the archive

- For newer mission the archives now contains all levels data
- The archive also include the software and calibration data
- This allows to :
 - Any point in time to reprocess the data
 - To apply new calibration and algorithms
- The most common data layout is a FITS event format
 - Each row contains attribute to a single event
- However some instruments have the data layout differently
 - Each row contains a spectrum or an array of spectra or an image integrated over a specific time interval
 - The different layout is driven typically by the instrument data mode (for example most of the RXTE and the Swift BAT survey data)

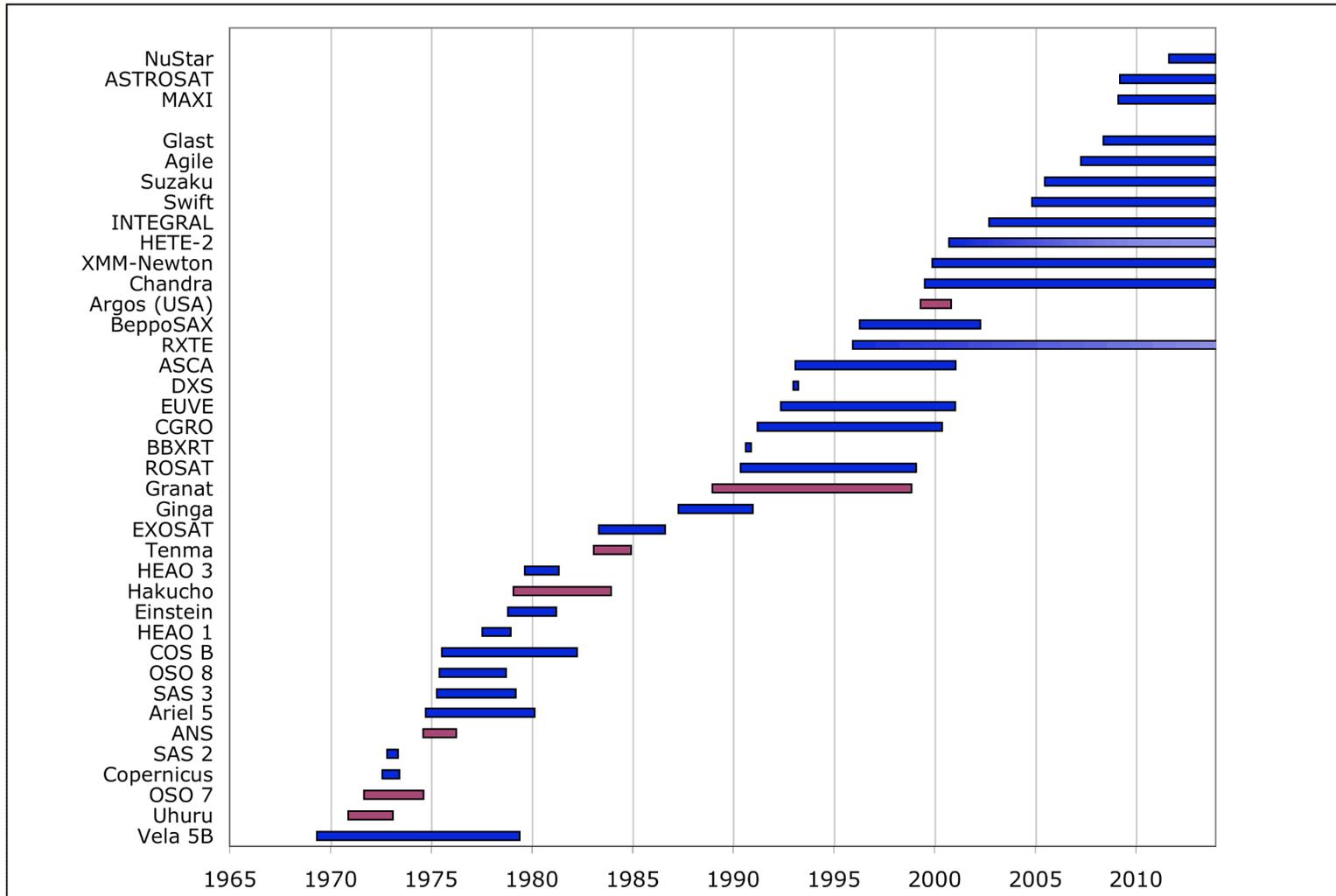
Archives

There are several high energy archive centers:

- HEASARC heasarc.gsfc.nasa.gov
- ASI Science Data Center www.asdc.asi.it
- Ledas ledas-www.star.le.ac.uk/
- Darts www.darts.isas.jaxa.jp

- Chandra Data Archive asc.harvard.edu/cda
- XMM Science Archive xmm.vilspa.esa.es/xsa/
- Integral isdc.unige.ch

High Energy data available in the Archives



Missions in blue are available in one or more archive centers listed in the previous slide
Nustar,MAXI and ASTROSAT are future mission that will be archived

HEASARC

- The physical archive includes :
 - Datasets from all the current and past X- and gamma-ray missions (also includes data located elsewhere, e.g. Chandra, XMM-Newton, BeppoSAX)
 - General stellar catalogs & access important Visir catalogs
- Provides data access via :
 - Browse: a search engine multi-mission and multi-catalog interface Web based to query database table and retrieve table
 - FTP and wget protocol for data retrieval
 - Batch interface for database query
 - Cross-correlation capabilities for database tables
- Provides multi-mission and mission specific software for the data in the archive , their calibration data information
- Include proposal preparation tools (Pimms Viewing) as well as general converter and calculation tools (e.g. Time, coordinate converter ; Nh calculator)

ASDC, LEDAS, DARTS

- ASDC in Italy is the main host for BeppoSAX and Agile data

Hosts also mirror copies of the Einstein, EXOSAT ASCA, ROSAT, Chandra, XMM, INTEGRAL, Swift and soon also GLAST and several source catalogs

- LEDAS at Leicester University in the UK has mirrors of the Swift, Chandra, ASCA, ROSAT archives and the XMM source catalog

- DARTS at ISAS in Japan host data from GINGA, ASCA, Swift SUZAKU

CALDB

One of the problems in X-ray and Gamma ray astronomy is the large number of files required to store all the calibration information for an instrument. Keeping track of these and making these available to the data analysis software is challenging.

- The CALDB is a directory structure and indexing system for FITS files which enables software to read the correct calibration data without needing to know the filename or directory. A benefit is that the calibration information is NOT embedded within the software.
- Users can either install a CALDB locally or for occasional use access the calibration files from the HEASARC site.
- CALDB is used for HEASARC archival and current missions and for Chandra but not (alas) for XMM.

SOFTWARE Packages

The data in the archive are accompanied by software packages that allow to analyze and manipulate the data. There are

HEASoft - heasarc.gsfc.nasa.gov/docs/software/lheasoft

CIAO - asc.harvard.edu/ciao

SAS - xmm.vilspa.esa.es/external/xmm_sw_cal/sas_frame.shtml

Integral - software.isdc.unige.ch

All these packages operate on FITS files with standard formats defined for each type of dataset.

It is possible to mix and match tools from different packages to perform specific tasks

HEASOFT

- Large and diverse package distributed by NASA Goddard.
- Run on several operating system platforms
- Can be downloaded as either source code or binaries. The latter is usually adequate unless you want to apply source patches.
- Contains generic multi-mission tools as well as packages for current missions RXTE, Swift, Suzaku and a number of archival missions.
- For help : heasarc.gsfc.nasa.gov/cgi-bin/ftoolshelp

HEASOFT SubPackages

- XANADU - High-level, multi-mission tasks for X-ray astronomical spectral, timing, and imaging data analysis
- FTOOLS - General and mission-specific tools for FITS files
- FITSIO - Core library responsible for reading/writing FITS files
- fv - General FITS file browser/editor/plotter with a graphical user interface (distributed with FTOOLS)
- XSTAR - Tool for calculating the physical conditions and emission spectra of photo-ionized gases

- PROFIT - Visualizing and modeling high-resolution spectra.

CIAO

- Written for Chandra by the CXC but contains many tools that can be used for other purposes.
- Run on several operating system platforms
- Software distribution available as source code and binary. The latter is recommended
- For help : asc.harvard.edu/helpdesk
- Includes :
 - Chandra-specific tools
 - Multi-mission DM (DataModel) tools
 - ds9 image display
 - ChIPS plotting package
 - Sherpa multidimensional fitting engine
 - GUIDE tool for identifying atomic features in spectra (uses the APED atomic physics database)

SAS & INTEGRAL

- SAS : Tools to be used on data from XMM-Newton.
 - Binary versions only are available on different operating system .
 - Requires the user to have installed : ds9, XPA, Grace, HEAsoft
 - For help :
xmm.vilspa.esa.es/external/xmm_user_support/xmmhelp_frame.shtml
- Integral : Tools to be used on data from INTEGRAL mission
 - Source code and Binary (on different operating system) version are available.
Recommended to install the binary
 - Requires the user to have installed ROOT , HEAsoft, DS9, wget
 - For help : inthelp@sciops.esa.int

HERA : environment to remote data analysis

- Run your analysis remotely on machines at GSFC.
(<http://heasarc.gsfc.nasa.gov/hera/startinghera.html>)
- Uses fv as a front-end. fv is a small stand-alone program that runs on Windows, Linux, Mac OS X.
(<http://heasarc.gsfc.nasa.gov/docs/software/ftools/fv>)
- Through Hera you can run programs from HEASoft, CIAO, and SAS without having to install any of them on your computer.
- Hera has high-speed access to the HEASARC archive.
- On Linux and Mac OS X systems there is a command line interface :
- `fv -r <HEASoft, CIAO or SAS command>`

SOFTWARE OVERVIEW

The path to data analysis involves steps that can be divided in the following classes :

- Calculating event attributes - sky position, PI, grade,...
 - Cleaning and filtering events
 - Extracting products - images, spectra, lightcurves,...
 - Creating observation calibration files
 - High-level analysis - source detection, spectral fitting
- NOTE : this general path is not always applicable to all data. Consideration such as large field of view or data format different from event may require different treatment

Filling event files

- The first step is to calculate various event attributes such as the sky position, the energy, and, for CCDs, the grade.
- Usually this will be performed automatically for you in pipeline processing using mission-dependent software that access the mission calibration data
- However, if calibration files change the data need to be reprocessed
 - Look for calibration updates
 - Use the mission dependent software to reprocess

Cleaning and filtering

- The next step is to remove unwanted events. The criteria depend :
 - On the mission characteristics such as orbital parameters
 - On the instrument characteristic such as specific housekeeping parameters (temperature voltages etc..)
 - On the event characteristics such as bad pixel
 - Times of particularly bad background.
- Additional selection includes only events within a particular region on the sky or on the detector.

There are many tools to select rows out of FITS files

= > but be a bit careful about which ones are used. Other software may need to know what selections were made.

For example time interval selection: from the only event left after the time selection, it is not possible to reconstruct the selection because data may include real gaps. The time intervals used in the selection need to be stored with the data

Cleaning and filtering II

- Selections are tracked by storing additional information in the file either in keywords or extra extensions.
- In CIAO this is done using the DataModel filtering. See `ahelp dmfiltering` for details.
- The SAS has either a command line option called `evselect` or a GUI called `xmmselect` (xmm.vilspa.esa.es/sas/current/doc/evselect).
- HEASoft has a program `xselect` which can be used to perform filtering (through the `filter` command). `xselect` actually runs an ftool called `extractor`, which is available for the sophisticated user (<http://heasarc.gsfc.nasa.gov/docs/software/lheasoft/ftools/xselect/xselect.html>).
- All these methods *should* be equivalent.

Extracting products

- Images, energy spectra, and lightcurves are histograms on event attributes
- General software (such as fselect or fhisto) allow to generate these files but in many cases subsequent software requires particular keywords or extensions so it is better to use standard tools.
 - The HEASoft program xselect has a command extract which can be used to make these products that support most of the mission.
 - evselect and xmmselect in SAS have options to create these products.
 - In CIAO dmextract is used to make energy spectra and lightcurves and dmcopy to make images.

Creating observation calibration files

- Before start the scientific analysis, observation dependent calibration files may need to be created. For example, analysis of images may need exposure maps and spectra require responses.
- These tools are always mission-specific since they depend on details of the instruments in use.
 - For Chandra look at the appropriate analysis threads in :
<http://asc.harvard.edu/ciao/threads/index.html>
 - For SAS a good place to start is the XMM ABC guide at :
<http://xmm.gsfc.nasa.gov/docs/xmm/abc/abc.html>
 - For many of the missions which the software is distributed within HEASoft (for example Swift and Suzaku) follow the instruction provided in the mission user guide (you can find that following mission links from <http://heasarc.gsfc.nasa.gov>)

Analysis of products

- When you have your spectrum, lightcurve or image and the relevant calibration file you can start the scientific analysis
- Source detection in images, fitting models to energy spectra, searching for periods in lightcurves,...
- The XANADU Package allows for mission independent analysis. This includes :
 - Imaging Analysis (via Ximage, detection routines also in Ciao)
 - Timing Analysis (via XRONOS)
 - Spectral Analysis (via Xspec , Sherpa also in Ciao)

Word of Advice

Although data and software are now more friendly :

Make sure you understand the instrument you are using

Check if the calibration is the latest and appropriate to your dataset

Make sure that the software gives you sensible results

Data centers do their best possible work to populate the archive
with high quality data but always ask questions

.... Do not use data, software and calibration as a black box